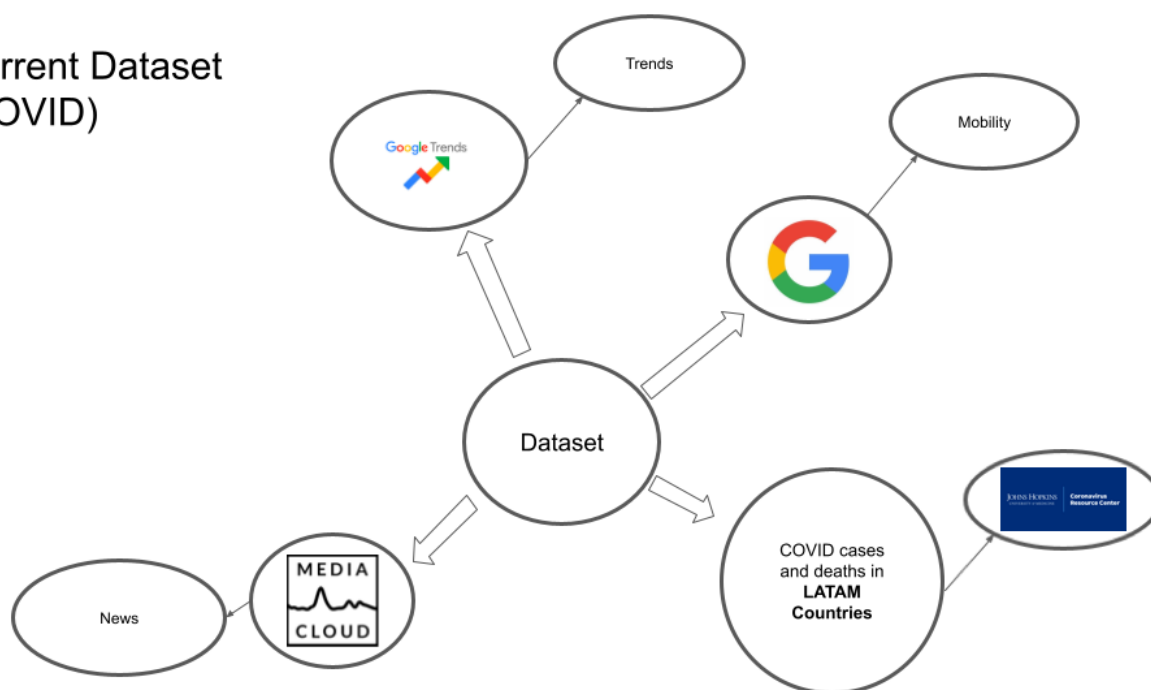


Infodemia y desinformación en la propagación de COVID-19 en LATAM.

Current Dataset
(COVID)



El dataset se compone de 3 fuentes de datos diferentes:

1. Datos de COVID-19 en LATAM, obtenidos de Our World in Data
2. Fuentes de información de internet, las cuales son:
 - 2.1. Media Cloud
 - 2.2. Google Trends

NOTA: Para mas información del dataset y códigos usados durante su generación, consultar el repositorio: https://github.com/MITCriticalData-Colombia/Covid_Media_Cloud

1. Datos de COVID-19 en LATAM

Los datos de COVID de LATAM se compone de 7 diferentes datasets:

- **covid_latam(raw).csv:** Contiene los datos de COVID para 32 países de LATAM. El dataset contiene las columnas:
 - 'iso_code': Código ISO del país de 3 caracteres.
 - 'continent': Continente de donde provienen los datos.

- 'location': País de donde provienen los datos.
- 'date': Fecha en la cual se registraron los datos.
- 'total_cases': Casos de covid totales (acomulativo) o casos probables.
- 'new_cases': Casos de covid nuevos presentados en el país y fecha establecidos.
- 'New_cases_smoothed': Nuevos casos confirmados haciendo una aproximación de la semana actual.
- 'Total_deaths': Total muertes atribuidas al covid-19.
- 'New_deaths': nuevas muertes atribuidas de covid-19
- 'New_deaths_smoothed': nuevas muertes de covid-19 haciendo una aproximación de la semana actual.
- 'Total_cases_per_million': Total de casos confirmados de COVID-19 por cada 1.000.000 de personas.
- 'New_cases_per_million': Nuevos casos confirmados de COVID-19 por cada 1.000.000 de personas.
- 'New_cases_smoothed_per_million': Nuevos casos confirmados de COVID-19 por cada 1.000.000 de personas haciendo una aproximación de la semana actual.
- 'Total_deaths_per_million': Total de muertes atribuidas a COVID-19 por cada 1.000.000 de personas
- 'New_deaths_per_million': Nuevas muertes atribuidas a COVID-19 por cada 1.000.000 de personas.
- 'New_deaths_smoothed_per_million': Nuevas muertes atribuidas a COVID-19 haciendo una aproximación de la semana actual por cada 1.000.000 personas.
- 'Reproduction_rate': Estimación en tiempo real de la tasa de reproducción efectiva (R) de COVID-19, es decir, esta medida es la rapidez con la se está propagando el virus en una población en un momento dado.
- 'Icu_patients': Número de pacientes con COVID-19 en unidades de cuidados intensivos (UCI) en un día determinado.
- 'Icu_patients_per_million': Número de pacientes con COVID-19 en unidades de cuidados intensivos (UCI) en un día determinado por cada 1.000.000 personas
- 'Hosp_patients': Número de pacientes con COVID-19 en el hospital en un día determinado.
- 'hosp_patients_per_million': Número de pacientes con COVID-19 en el hospital en un día determinado por cada 1.000.000 personas.
- 'Weekly_icu_admissions': Número de pacientes con COVID-19 ingresados recientemente en unidades de cuidados intensivos (UCI) en una semana determinada.
- 'Weekly_icu_admissions_per_million': Número de pacientes con COVID-19 ingresados recientemente en unidades de cuidados intensivos (UCI) en una semana determinada por cada 1000000 personas.
- 'Weekly_hosp_admissions': Número de pacientes con COVID-19 recién ingresados en hospitales en una semana determinada.
- 'Weekly_hosp_admissions_per_million': Número de pacientes con COVID-19 recién ingresados en hospitales en una semana determinada por cada 1.000.000 personas
- 'total_tests': Pruebas totales para COVID-19

- 'New_tests': Nuevas pruebas para COVID-19 (solo calculadas para días consecutivos).
- 'Total_tests_per_thousand': Pruebas totales para COVID-19 por cada 1,000 personas.
- 'New_tests_per_thousand': Nuevas pruebas para COVID-19 por cada 1,000 personas.
- 'new_tests_smoothed': Nuevas pruebas para COVID-19 haciendo una aproximación de la semana actual. Para los países que no informan los datos de las pruebas a diario, asumimos que las pruebas cambiaron por igual a diario y se promedian en una ventana móvil de 7 días.
- 'New_tests_smoothed_per_thousand': Nuevas pruebas para COVID-19 haciendo una aproximación de la semana actual por cada 1,000 personas
- 'positive_rate': La proporción de pruebas de COVID-19 que son positivas, dada como un promedio móvil de 7 días (este es el inverso de tests_per_case).
- 'tests_per_case': Pruebas realizadas por nuevo caso confirmado de COVID-19, dadas como un promedio móvil de 7 días (este es el inverso positive_rate)
- 'tests_units': Unidades utilizadas por la ubicación para informar sus datos de prueba.
- 'Total_vaccinations': Número total de dosis de vacunación COVID-19 administradas.
- 'people_vaccinated': Número total de personas que recibieron al menos una dosis de vacuna.
- 'People_fully_vaccinated': Número total de personas que recibieron todas las dosis prescritas por el protocolo de vacunación inicial.
- 'Total_boosters': Número total de dosis de refuerzo de vacunación COVID-19 administradas.
- 'new_vaccinations': Nuevas dosis de vacunación COVID-19 administradas (solo calculadas para días consecutivos).
- 'New_vaccinations_smoothed': Nuevas dosis de vacunación COVID-19 administradas, dadas como un promedio móvil de 7 días.
- 'total_vaccinations_per_hundred': Número total de dosis de vacunación COVID-19 administradas por cada 100 personas en la población total.
- 'People_vaccinated_per_hundred': Número total de personas que recibieron al menos una dosis de vacuna por cada 100 personas en la población total.
- 'people_fully_vaccinated_per_hundred': Número total de personas que recibieron todas las dosis prescritas por el protocolo de vacunación inicial por cada 100 personas en la población total.
- 'Total_boosters_per_hundred': Número total de dosis de refuerzo de vacunación contra la COVID-19 administradas por cada 100 personas en la población total.
- 'New_vaccinations_smoothed_per_million': Nuevas dosis de vacunación COVID-19 administradas (7 días alisados) por cada 1.000.000 personas en la población total.
- 'New_people_vaccinated_smoothed': Número diario de personas que reciben su primera dosis de vacuna (7 días alisados) .
- 'new_people_vaccinated_smoothed_per_hundred': Número diario de personas que reciben su primera dosis de vacuna (suavizado de 7 días) por cada 100 personas en la población total.

- 'Stringency_index': Índice de rigurosidad de la respuesta del gobierno: medida compuesta basada en 9 indicadores de respuesta que incluyen cierres de escuelas, cierres de lugares de trabajo y prohibiciones de viaje, reescalado a un valor de 0 a 100 (100 = respuesta más estricta)
- 'Population': Población (últimos valores disponibles).
- 'population_density': Número de personas dividido por superficie terrestre, medido en kilómetros cuadrados, año más reciente disponible
- 'median_age': Edad media de la población, proyección de la ONU para 2020
- 'Aged_65_older': Proporción de la población de 65 años o más, año más reciente disponible
- 'aged_70_older': Porcentaje de la población de 70 años o más en 2015
- 'gdp_per_capita': Producto interno bruto a paridad del poder adquisitivo (dólares internacionales constantes de 2011), año más reciente disponible.
- 'Extreme_poverty': Proporción de la población que vive en la pobreza extrema, año más reciente disponible desde 2010.
- 'cardiovasc_death_rate': Tasa de mortalidad por enfermedad cardiovascular en 2017 (número anual de muertes por cada 100.000 personas).
- 'diabetes_prevalence': Prevalencia de diabetes (% de la población de 20 a 79 años) en 2017.
- 'Female_smokers': Proporción de mujeres que fuman, año más reciente disponible.
- 'male_smokers': Porcentaje de hombres que fuman, año más reciente disponible.
- 'Handwashing_facilities': Proporción de la población con instalaciones básicas para lavarse las manos en las instalaciones, año más reciente disponible.
- 'Hospital_beds_per_thousand': Camas de hospital por cada 1000 personas, año más reciente disponible desde 2010.
- 'Life_expectancy': Esperanza de vida al nacer en 2019.
- 'Human_development_index': Un índice compuesto que mide el logro promedio en tres dimensiones básicas del desarrollo humano: una vida larga y saludable, conocimiento y un nivel de vida digno.
- 'excess_mortality_cumulative_absolute': Diferencia acumulada entre el número informado de muertes desde el 1 de enero de 2020 y el número proyectado de muertes para el mismo período con base en años anteriores.
- 'Excess_mortality_cumulative': Diferencia porcentual entre el número acumulado de muertes desde el 1 de enero de 2020 y las muertes acumuladas proyectadas para el mismo período con base en años anteriores.
- 'excess_mortality': Diferencia porcentual entre el número informado de muertes semanales o mensuales en 2020-2021 y el número proyectado de muertes para el mismo período con base en años anteriores.
- 'Excess_mortality_cumulative_per_million': Diferencia acumulada entre el número informado de muertes desde el 1 de enero de 2020 y el número proyectado de muertes para el mismo período, con base en años anteriores, por millón de personas.

Nota: Para más detalles sobre el cálculo del exceso de mortalidad para los diferentes países, por favor revisar la siguiente información: <https://elifesciences.org/articles/69336>

- **Daily_covid_cases_latam.csv:** Casos de COVID diarios por millón de habitantes para 32 países de LATAM.
 - La primera columna 'location' contiene el nombre del país donde se presentaron los casos.
 - Las siguientes columnas contienen el número de casos por millón de habitantes presentados en ese día en ese país.
- **Daily_covid_deaths_latam.csv:** Muertes confirmadas por COVID diarias por millón de habitantes para 32 países de LATAM.
 - La primera columna 'location' contiene el nombre del país donde se presentaron las muertes.
 - Las siguientes columnas contienen el número de muertes confirmadas por millón de habitantes presentadas en ese día en ese país.
- **daily_covid_vaccinations_latam.csv:** Vacunaciones diarias confirmadas de COVID por cada 100 habitantes para 32 países de LATAM (de forma acumulada).
 - La primera columna 'location' contiene el nombre del país donde se presentaron las muertes.
 - Las siguientes columnas contienen el número de vacunas confirmadas por cada 100 habitantes aplicadas hasta ese día en ese país.
- **Weekly_covid_cases_latam.csv:** Casos de COVID semanales usando la semana epidemiológica por millón de habitantes para 32 países de LATAM.
 - La primera columna 'location' contiene el nombre del país donde se presentaron los casos.
 - Las siguientes columnas contienen la suma de casos por millón de habitantes presentados en ese país durante esa semana epidemiológica. El formato de la semana epidemiológica es:
 - 4 primeros dígitos correspondientes al año, seguidos de 2 dígitos correspondientes a la semana epidemiológica.
- **Weekly_covid_deaths_latam.csv:** Muertes confirmadas por COVID semanales usando la semana epidemiológica por millón de habitantes para 32 países de LATAM.
 - La primera columna 'location' contiene el nombre del país donde se presentaron las muertes.
 - Las siguientes columnas contienen la suma de las muertes por millón de habitantes presentadas en ese país durante esa semana epidemiológica. El formato de la semana epidemiológica es:
 - 4 primeros dígitos correspondientes al año, seguidos de 2 dígitos correspondientes a la semana epidemiológica.
- **weekly_covid_vaccinations_latam.csv:** Vacunas semanales confirmadas de COVID aplicadas utilizando la semana epidemiológica por cada 100 habitantes para 32 países de LATAM.

- La primera columna 'location' contiene el nombre del país donde se presentaron las vacunaciones.
- Las siguientes columnas contienen la suma del total de vacunaciones por cada 100 habitantes presentadas en ese país hasta esa semana epidemiológica. El formato de la semana epidemiológica es:
 - 4 primeros dígitos correspondientes al año, seguidos de 2 dígitos correspondientes a la semana epidemiológica.

Nota: Para los archivos que tienen información semanal, cada columna representa la semana epidemiológica en el formato (AñoSemana). Puedes usar la librería de Python [epiweek](#) para convertir las columnas a un formato de fecha si lo consideras necesario.

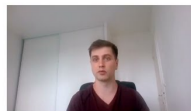
2. Internet sources:

2.1. Media Cloud

El conjunto de datos de Media Cloud se extrajo con la herramienta Explorer de Media Cloud. Algunos ejemplos del uso de la herramienta se pueden consultar en los videos:

1. [Media Cloud - Topic Mapper](#):

Media Cloud Topic Mapper



Adrien Carrel



2. [Media Cloud - Explorer](#):

Media Cloud - Explorer

How it works:



Powerful query tool to search through diverse digital news media sources



Allows search by multiple languages (English, Spanish, Arabic, ...)



Includes temporal information which enables the observation of trends



Features various representations of how topics are covered (word-clouds, bi-grams, embeddings)



Provide additional information about stories to better understand the scope of your topic of interest



Los datasets que encontraremos en esta sección son:

- **covid-19_attention_latam.csv**: Atención a la palabra COVID-19 en los países de LATAM.
 - La primera columna 'date' contiene la fecha en la cual se están analizando la atención a la palabra.
 - Cada país tiene las columnas para la palabra "covid-19" en cada país:
 - "count": Representa el número de noticias en media cloud que mencionan nuestro tópico.
 - "total_count": Representa el número total de noticias en media cloud que fueron publicadas.
 - y "ratio": Contiene la porción de noticias en media cloud que contienen el tópico de interés
 - La columna se presenta en el formato:
 - count_NombreDelPais, total_count_NombreDelPais, ratio_NombreDelPais
- **covid_vaccination_attention_latam.csv**: Atención a la palabra vacunación de COVID en los países de LATAM.
 - La primera columna 'date' contiene la fecha en la cual se están analizando la atención a la palabra.
 - Cada país tiene las columnas para la palabra "vacunación de COVID" en cada país:
 - "count": Representa el número de noticias en media cloud que mencionan nuestro tópico.
 - "total_count": Representa el número total de noticias en media cloud que fueron publicadas.
 - y "ratio": Contiene la porción de noticias en media cloud que contienen el tópico de interés
 - La columna se presenta en el formato:
 - count_NombreDelPais, total_count_NombreDelPais, ratio_NombreDelPais.

2.2. Google Trends

El conjunto de datos de Google Trends se extrajo con la herramienta Online de google trends. Algunos ejemplos del uso de la herramienta se pueden consultar en el video:

2. [Google Trends tutorial](#):

Google Trends

How it works:

1. Compares the number of times the keyword appeared in searches with the total number of searches in a specific region and time.
2. The results of the comparison are ranked between 0 and 100 representing the popularity of the topic in a certain time and location.

Examples:

1. Forecasting Influenza using google trends [1].
2. Early warning COVID-19 monitor [2].
3. COVID-19 lockdowns and well-being [3].



[1] - Lu F, Hou S, Baltrusaitis K, Shah M, Leskovec J, Sosic R, Hawkins J, Brownstein J, Conidi G, Gunn J, Gray J, Zink A, Santillana M. Accurate Influenza Monitoring and Forecasting Using Novel Internet Data Streams: A Case Study in the Boston Metropolis. JMIR Public Health Surveill 2018;4(1):e4. DOI: 10.2196/publichealth.8950
[2] - Kogan NE, Clemente L, Liautaud P, Kaashoek J, Link NB, Nguyen AT, Lu FS, Huybers P, Resch B, Havas C, Petutschnig A, Davis J, Chinazzi M, Mustafa B, Hanage WP, Vespignani A, Santillana M. An early warning approach to monitor COVID-19 activity with multiple digital traces in near real time. Sci Adv. 2021 Mar 5;7(10):eabd6989. doi: 10.1126/sciadv.abd6989. PMID: 33674304; PMCID: PMC7935356
[3]- Brodeur A, Andrew EC, Sarah F, Nattavudh P. (2020). COVID-19, lockdowns and well-being: Evidence from Google Trends. Journal of Public Economics. 193. 10.1016/j.jpubeco.2020.104346.

El dataset que encontraremos en esta sección es:

- **covid-19_Symptoms_latam.csv:** Tendencias de algunos síntomas de Covid-19 en países de LATAM.
 - La primera columna 'Week' contiene la fecha en la cual se están analizando las tendencias de los síntomas en LATAM.
 - Cada país tiene las columnas Fever (Fiebre), Headache (Dolor de Cabeza) y Cough (Tos), que representan algunos de los síntomas de covid-19 en cada país. La columna se presenta en el formato:
 - Fever_NombreDelPais, Headache_NombreDelPais, Cough_NombreDel-Pais